Structure of the P22 \( \text{att} \) Site

CONSERVATION AND DIVERGENCE IN THE \( \lambda \) MOTIF OF RECOMBINOGENIC COMPLEXES*

(Received for publication, March 17, 1994, and in revised form, June 2, 1994)

Lynda Smith-Mungo, Iris T. Chan‡, and Arthur Landy§
From the Division of Biology and Medicine, Brown University, Providence, Rhode Island 02912

We have defined the bacterial and viral DNA targets (\( \text{att} \) sites) of P22 site-specific recombination and characterized their interaction with integrase (Int) protein. The bacterial DNA target, \( \text{attB} \), is approximately 27 base pairs and consists of two core type Int binding sites as inverted repeats. The top and bottom Int cleavage sites fall within the core type \( \text{att} \) binding sites and are separated by a 7-base pair overlap region. A similar core region is found in the viral DNA target, \( \text{attP} \), which is approximately 260 base pairs long and contains two IHF binding sites and five arm type binding sites for Int. The results suggest that P22 Int, like \( \lambda \) Int, is a heterobivalent DNA-binding protein that is capable of forming complex higher order structures with recombinogenic function. Although P22 and \( \lambda \) recombination involve very similar multiprotein interactions and core region structures, there are significant differences in the arrangements of distal protein binding sites. These differences are discussed in terms of the possible flexibility of the Int protein and the specificity with which the higher order complexes assemble and/or function.

The site-specific recombination pathway of the *Salmonella typhimurium* phage P22 is responsible for the integration and excision of the viral genome into and out of the host chromosome during the transition between lytic and lysogenic states (for reviews see Refs. 1 and 2). The Int\(^1\) protein catalyzing this reaction belongs to the 30-member Int family of recombinases that carry out conservative site-specific recombination in the genomes of both prokaryotes and eukaryotes and proceeds via a reciprocal site-specific recombination, as first proposed by Campbell (3). The Int family is defined by an active site tyrosine and a conserved Arg-His-Arg motif that is thought to be important in the chemistry of DNA strand cleavage and ligation (4, 5). These reactions, which do not require any high energy cofactors, utilize a covalent phosphotyrosine intermediate to carry out DNA strand cleavage and ligation. DNA:DNA homology between recombining partners is required within a short overlap region that is delimited by the recombinase cleavage sites, which are staggered by 6–8 bp, depending upon the recombination system. The cleavage sites fall near the edges of recombinase binding sites that occur as inverted repeats flanking the overlap region. The segment of DNA comprising the overlap region and flanking recombinase binding sites is approximately 25–30 bp and is referred to as the core region (for recent reviews see Refs. 6–10).

The Int family can be divided into two groups on the basis of the complexity of the recombination pathways. Type I proteins are exemplified by the Cre and FLP systems: they do not require accessory proteins; involve two similar, relatively short, DNA segments as recombination partners (\( \text{att} \) sites); and consist essentially of a core region (11, 12). Type II proteins were first characterized in the recombination pathway of bacteriophage \( \lambda \), which involves dissimilar recombination partners that have supplemented the basic core region with multiple binding sites for the recombinase and the required accessory DNA-bending proteins. Integrative recombination between phage (\( \text{attP} \)) and Escherichia coli DNA (\( \text{attB} \)) requires the phage-encoded Int and the host-encoded IHF proteins. The products of this reaction, the prophage sites \( \text{attL} \) and \( \text{attR} \), are the partners for excisive recombination, which additionally requires the phage-encoded Xis protein and is stimulated by the host-encoded FIS protein (6, 7). Two recombination pathways, those of phage 434 and HK022, have been shown to be virtually identical to \( \lambda \) in the structure of their recombination sites (13, 14), whereas many other pathways that are quite different from \( \lambda \) at the DNA sequence level are known, or likely to be, type II on the basis of extended \( \text{att} \) sites and/or a requirement for accessory proteins. The best studied of these pathways are those of the *E. coli* phage P2, the *Hemophilus influenzae* phage, HP1, and the mycobacteriophage, L5 (15–19). To gain some insights into the range of type II recombinase structural motifs and how they might be utilized in different ways we chose to study P22 Int.

The recombination pathway of bacteriophage P22 is similar to \( \lambda \) in its use of a virally encoded Xis protein and a cellularly encoded IHF protein (1, 2, 20–23). However, like many other Int family members, the four \( \text{att} \) sites of P22 share a larger region of homology than is observed in \( \lambda \), 46 bp versus 15 (22, 24).

In this report we show that in addition to their similar multiprotein interactions, P22 shares with \( \lambda \) and P2 the feature of a heterobivalent Int protein and \( \text{att} \) sites with nearly identical core regions. However, despite all of the similarities between the \( \lambda \) and P22 pathways, the arrangement, spacing, and orientation of their Int binding sites are surprisingly different, thus raising a number of interesting questions about the formation, structure, and function of the multiprotein higher order DNA complexes that govern this class of recombination reactions.

EXPERIMENTAL PROCEDURES

Plasmids—The following plasmids were used in the course of this work. pJLR1 (Int\(^1\)) (made by J. Ross) was derived by the deletion of a NneI-StuI Int gene containing fragment from pJL110 (pJL110 contains a P22 \( \text{attP} \) fragment from BamHI (0.481)-AvaI (0.568) cloned into the BamHI-PvuII backbone of pH322; see Ref. 22. pMC2 is a smaller \( \text{attP} \)-

---

* This work was supported by National Institutes of Health Grants AI 13544 and GM 33926. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present address: Howard Hughes Medical Institute, Dept. of Medicine, Genetics and Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110.
§ To whom correspondence should be addressed. Tel.: 401-863-1658; Fax: 401-663-2421.

\(^1\) The abbreviations used are: Int, integrase; bp, base pair(s); PCR, polymerase chain reaction; NCS, neocarzinostatin.
containing fragment derived from JJR1 by the deletion of a 900-bp HindIII fragment. pMR107 contains a 1,350-bp HindIII P22 attB fragment cloned into pBR322 backbone (gift of M. Riley (21)). The plasmids containing truncations of the attP and attB of P22 were prepared by a strategy of primer-directed mutagenesis using the polymerase chain reaction (PCR) and oligonucleotide primers synthesized in a Biosearch 8600 DNA synthesizer or Applied Biosystems 394 DNA/RNA Synthesizer. The oligonucleotide primers for the P' arm and B' deletions anneal to the top strand of their respective att sequences and consist of 2 nonspecific bp of DNA (AT) and a BamH I restriction enzyme cloning site located 8-20 bases of complementary att sequences. The primers for the P and B arm deletions anneal to the bottom strand of their respective att sequences and consist of 2 nonspecific bp of DNA (GG) and a HindIII restriction enzyme site followed by 18-20 bases of complementary att sequence. The primers used to construct pIC18 and pIC19 contained an additional 8 bp of nonspecific DNA of random sequence between the BamH I site and the complementary att sequence. The end points of all of the truncated att sequences are numbered according to Ref. 22 and are listed in Table I. The PCR was performed in a Perkin-Elmer Thermal Cycler in 100-μl reaction volumes containing 5 ng of template DNA, 20 pmol of each primer set, a 200 μM concentration of each dNTP, 10.0 mM Tris-HCl, pH 8.3, 50.0 mM KCl, 1.5 mM MgCl₂, 0.1% (w/v) gelatin, and 2.5 units of AmpliTaq™ DNA polymerase. The P arm deletion plasmids were generated using HindIII-digested pIC2 as the template for PCR. The P' arm deletion plasmids were generated using Thd 111-digested pIC4 or pIC5 as the template for PCR. In these PCR reactions, both the vector DNA and attP DNA were amplified. The attP-containing PCR fragments were electrophoresed on a 1.2% low melting agarose gel and purified by the method of GeneClean™ (Bio 101, Inc.). The PCR fragments containing 7 arm deletions were ligated following HindIII digestion to generate pIC3, 4, 5, 6, and 7; the PCR fragments containing B arm deletions were ligated following BamH I digestion to generate pIC8, 9, 10, 11, and 12. The attP deletion plasmids were transformed into E. coli strain HB101 and sequenced using the Sequenase™ kit (U. S. Biochemical Corp.). The B and B' arm deletions were generated using pMR107 as the template for PCR reactions in which only attP sequences were amplified. These small, linear PCR products were used directly in the recombination assays. The fragments containing the attB deletions were also subcloned into HindIII-BamH I-digested pBR322 to generate pIC13, 14, 15, 16, 17, 18, and 19. The attP and attB deletion mutants and their 5' and 3' boundaries are summarized in Table I.

**Recombination Assays and Proteins**—Integrative recombination assays were carried out with 0.3 μg each of supercoiled attP and linear attB substrates in 20 μl of recombination buffer (25 mM Tris-HCl, pH 7.9, 75 mM NaCl, 6 mM spermidine, 5 mM EDTA, 2 mM dithiothreitol, 0.5 mg/ml bovine serum albumin) with 2-3 units of purified E. coli IHF (25). Two-fold dilutions of purified P22 Int were (4, 2, and 1 units) were added to each reaction and incubated for 3-4 h at 25 °C. The minimum amount of P22 Int protein which gives maximal recombination under these conditions is defined as 1 unit. The purification of P22 Int will be described elsewhere. The reactions were stopped with 2 μl of stop solution (1% sodium dodecyl sulfate, 10% Ficoll, 0.2% bromphenol blue) and electrophoresed on 1.2% agarose gels at 50 V for 16 h or 100 V for 4 h. The gels were stained with ethidium bromide and visualized under UV light. Recombination efficiency was scored as + + (equal to or greater than 50% recombination), + (less than 50% recombination), or (no detectable recombination).

2 M. Horvath, J. Rich, and C. Leng, unpublished results.

### Table I

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>B arm</th>
<th>B' arm</th>
<th>Plasmid</th>
<th>P arm</th>
<th>P' arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMR107</td>
<td>-235</td>
<td>+84</td>
<td>pIC2</td>
<td>-773</td>
<td>+180</td>
</tr>
<tr>
<td>pIC13</td>
<td>-25</td>
<td>+50</td>
<td>pIC3</td>
<td>-222</td>
<td>+180</td>
</tr>
<tr>
<td>pIC14</td>
<td>-18</td>
<td>+50</td>
<td>pIC4</td>
<td>-201</td>
<td>+180</td>
</tr>
<tr>
<td>pIC15</td>
<td>-13</td>
<td>+50</td>
<td>pIC5</td>
<td>-156</td>
<td>+180</td>
</tr>
<tr>
<td>pIC16</td>
<td>-25</td>
<td>+22</td>
<td>pIC6</td>
<td>-106</td>
<td>+180</td>
</tr>
<tr>
<td>pIC17</td>
<td>-17</td>
<td>+5</td>
<td>pIC7</td>
<td>-67</td>
<td>+180</td>
</tr>
<tr>
<td>pIC18</td>
<td>-25</td>
<td>-7</td>
<td>pIC8</td>
<td>-201</td>
<td>+63</td>
</tr>
<tr>
<td>pIC19</td>
<td>-25</td>
<td>-7</td>
<td>pIC9</td>
<td>-201</td>
<td>+90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pIC10</td>
<td>-201</td>
<td>+138</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pIC11</td>
<td>-156</td>
<td>+180</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pIC12</td>
<td>-156</td>
<td>+160</td>
</tr>
</tbody>
</table>

### Table II

<table>
<thead>
<tr>
<th>Fragments at endpoint</th>
<th>Recombination efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-235</td>
<td>++</td>
</tr>
<tr>
<td>-25</td>
<td>++</td>
</tr>
<tr>
<td>-13</td>
<td>++</td>
</tr>
<tr>
<td>-25</td>
<td>++</td>
</tr>
<tr>
<td>-25</td>
<td>++</td>
</tr>
<tr>
<td>-25</td>
<td>++</td>
</tr>
<tr>
<td>-25</td>
<td>++</td>
</tr>
</tbody>
</table>

*Recombination efficiency was judged by the amount of product formed at various Int concentrations as visualized by ethidium bromide staining. +, greater than or equal to 50% recombination; ++, less than 50% recombination; −, no detectable recombination.*
to near homogeneity as judged by silver staining (23). An in vitro recombination reaction with supercoiled attP and linear attB, P22 Int, and E. coli IHF was carried out as described under "Experimental Procedures." The linear recombination product migrates near the top of an agarose gel, just above the lane corresponding to nicked circular plasmid (Fig. 1, lanes 1–4). To determine the minimum DNA sequence required to define attB, primers complementary to the attB sequence were synthesized. Truncated fragments whose end points correspond to the outer boundaries of the primers were then synthesized by PCR. (The boundaries of these fragments are listed in Table I.) Fragments containing deletions from the right or left sides of attB were then used as substrates in a series of in vitro recombination assays. When an attB encompassing residues −25 to +50 is used in the reaction, a recombinant product (recomb. (trun.)) is formed which migrates faster than the attB substrate in lanes 1–4 (Fig. 1, lanes 5–8). Further truncation of the attB site to −18 abolishes the ability of this fragment to recombine (Fig. 1, lanes 9–12), suggesting that some residues from −25 to −19 are critical to the reaction. Residue −25 was then chosen as the boundary of the B arm, and successive truncations of the B’ arm were then tested in an identical manner. Fragments containing residues −25 to +20 and −25 to +1 were able to recombine; however, further deletion of the attB to −7 resulted in undetectable levels of recombination indicating the importance of residues −7 to +1. A summary of all of the fragments synthesized and tested in identical assays and the resulting recombination efficiencies is provided in Table II. These data indicate that attB is contained within a 27-bp fragment that encompasses residues −25 to +1. This sequence is also equivalent to the core region, as discussed below.

Similar studies were performed on the attP-containing plasmid to define the functional unit of integrative recombination for this partner (see "Experimental Procedures"). As shown in Table III, the recombination efficiency when the P arm is truncated from residue −773 to −201 shows no detectable change. However, the truncation to residue −153 decreases recombination efficiency, whereas further truncation to residue −104 completely abolishes the formation of product.

In another series of experiments, an attP plasmid with a P arm boundary at −201 was used as the template for PCR truncation of the P’ arm. When the resulting supercoiled attP plasmids were tested as recombination partners it was found that truncation is decreased when the P’ arm is shortened from +50 to +90 and completely abolished after truncation to residue +63. Footprinting experiments (see below) indicated an arm type Int binding site at residues 87–96. The primer used to make the +90 truncation reconstituted an arm type Int binding site (TGGGGGATCC) with only three mismatches from the consensus sequence. Because the affinity of Int for this junction sequence was not known we created additional plasmids in which the attP extended to residue +86 (completely removing the P’2 site) or +106 (completely restoring the P’2 site). Recombination

**TABLE III**

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>End point (P/P’ arm)</th>
<th>Recombination efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>pIC2</td>
<td>−773/+180</td>
<td>++</td>
</tr>
<tr>
<td>pIC3</td>
<td>−222/+180</td>
<td>++</td>
</tr>
<tr>
<td>pIC4</td>
<td>−201/+180</td>
<td>++</td>
</tr>
<tr>
<td>pIC5</td>
<td>−158/+180</td>
<td>+</td>
</tr>
<tr>
<td>pIC6</td>
<td>−106/+180</td>
<td>−</td>
</tr>
<tr>
<td>pIC7</td>
<td>−57/+180</td>
<td>−</td>
</tr>
<tr>
<td>pIC10</td>
<td>−201/+138</td>
<td>++</td>
</tr>
<tr>
<td>pIC9</td>
<td>−201/+90</td>
<td>+</td>
</tr>
<tr>
<td>pIC8</td>
<td>−201/+63</td>
<td>−</td>
</tr>
<tr>
<td>pIC11</td>
<td>−156/+86</td>
<td>−</td>
</tr>
<tr>
<td>pIC12</td>
<td>−156/+106</td>
<td>+</td>
</tr>
</tbody>
</table>

* Recombination efficiency was judged by the amount of product formed at various Int concentrations as visualized by ethidium bromide staining: ++, greater than or equal to 50% recombination; +, less than 50% recombination; −, no detectable recombination.

**FIG. 1.** In vitro recombination between wild type P22 attP and different attB substrates. Wild type P22 supercoiled attP was incubated under standard recombination conditions with increasing dilutions of purified P22 Int and either wild type or truncated attB substrates formed by PCR. The reactions were electrophoresed on 1.2% agarose gels and the products visualized by ethidium bromide staining. Full-length linearized substrates formed by PCR. The reactions were electrophoresed on 1.2% agarose gels and the products visualized by ethidium bromide staining. Full-length linearized attB (lanes 1–4), truncated attB encompassing residues −25 to +50 (lanes 5–8), and truncated attB encompassing residues −18 to +50 (lanes 9–12) were used. In each assay purified Int was sequentially diluted prior to addition to a final concentration product; recomb., wild type recombination product; attP (rel.), relaxed attP; attB (lin.), linearized wild type attB; recomb. (trun.), truncated recombination product; s.c. attP, supercoiled wild type attP; attB (trun.), truncated attB.

**FIG. 2.** P22 Int cleavage of bottom strand-labeled attP. An XmaI-MaeIII fragment of the P22 attP encompassing the putative overlap region was 3' end labeled on the bottom strand and gel purified. Following incubation with different concentrations of integrase, the reactions were electrophoresed on a sequencing gel and products visualized by autoradiography. The numbers on the right indicate the base number within the wild type attP as determined by a Maxam and Gilbert A+G sequencing lane. Lane headings are as follows: NP, no protein added; 8, 4, 2, 1, number of units of Int added per reaction; Seq., sequence.
was observed only when the P'2 site was intact (Table III).

Localization of the P22 Int Cleavage Sites—To determine the P22 Int cleavage sites we used the approach first described by Craig and Nash (28). Briefly, a bottom strand 3' end-labeled attP-containing fragment was prepared and incubated with P22 Int in the absence of partner, which should allow DNA cleavage but not formation of recombinant. When the products were analyzed on a sequencing gel, an Int-dependent band that comigrates approximately with residue −10 on the sequencing ladder is evident (Fig. 2, lanes 4–5). By analogy with all other members of the Int family, which cleave DNA by forming a covalent 3' phosphotyrosine bond, the observed fragment is proposed to extend from an Int-generated 5' hydroxyl to the 32P-labeled 3' end. The Int-generated fragment has one less phosphate than a fragment with the same number of bases in the sequencing ladder. Although this can result in a mobility that is approximately one-half step slower than the corresponding band in the sequencing ladder, such intermediate mobility was not resolved in our gels (28). The low yield of the cleavage product indicates that, like other members of the Int family, the equilibrium in the P22 cleavage- ligation reaction lies far toward the side of ligation. An even lower yield of Int cleavage product was observed when similar experiments were carried out with the 32P 3' end label on the top strand. The Int cleavage product of the top strand, which was difficult to reproduce photographically, comigrates with residue −18 of the analogous sequence ladder (data not shown). The results indicate that P22 Int cleaves between residues −17 and −18 on the top strand and between −10 and −11 on the bottom strand.

Core Type Int Binding——The interaction of P22 Int with attB was investigated by DNase and NCS protection experiments. An attB-containing restriction fragment with a single 32P 5'-terminal label in the top strand was incubated with the indicated proteins and then digested with DNase (Fig. 3A) or NCS (Fig. 3B). In the DNase footprint of the attB we observed an area of protection by P22 Int between bases −30 and +4 (Fig. 3A, lanes 1–4). The NCS footprint of attB is less dramatic in its protection pattern as the cleavage of bases in the attB by this reagent appear less intense in the region of interest (Fig. 3B, lanes 1–3, 5–7). In the NCS footprint most of the core protection is lost by the final Int dilution. Nevertheless, at the higher concentration of P22 Int, residues +1, −5 to −8, −12, −19, and −23 to −28 are seen to be protected by P22 Int. The results of the two types of footprinting analysis are consistent with each other as all residues protected against NCS cleavage fall within the region of DNA protected by DNase.

In the numbering system of P22 attP, the zero coordinate was assigned to the center of an imperfect inverted repeat that was itself approximately centered between the two IHF sites (22). This inverted repeat was considered to be a possible candidate for the core type Int binding sites; however, the attB footprinting data indicate that the majority of the protection and enhancements occurs from bases −32 to +4 (Fig. 3A, lanes 1 and 4; Fig. 3B, lanes 1–3 and 5–7). To investigate this result further,
P22 Int binding to the phage att site was analyzed by DNase and NCS footprinting.

The pattern of attP protection in the presence of P22 Int is more complex than that obtained with the attB. We will first address the protection of residues located between the IHF binding sites. The protection includes residues -26 to +4 in the DNase footprint (Fig. 4A, lanes 1–3 and 7–9) and residues -24 to +2 in the NCS footprint (Fig. 4B, lanes 1–3 and 7–9). Finally, in two separate DMS footprinting experiments Int protection of attP residues -12, -9, -6, -5, -4, and +2 on the top strand and residues -1, -2, and -8 on the bottom strand was observed (for summary see Fig. 5, A and B).

In summary, the footprinting results with the attB and attP attachment sites are consistent with each other and indicate that the core type Int binding sites are localized in the area of -30 to +4. This is at the left boundary of the region of homology between the attP and attB DNA, which encompasses 46 bp (-18 to +27) (22). The only inverted repeat within the Int-protected core region which aligns reasonably with the Int cleavage sites and/or falls within the functional boundaries defined by the attB truncation experiments is the consensus sequence element CTTCGCATT (-24 to -4). The sequences of the four core type sites and the resulting consensus sequence are summarized in Fig. 6A.

Arm Type Binding Sites in the Phage Attachment Site—A number of areas of protection were visible by DNase footprinting of P22 Int on the P22 viral attachment site. These include bases -150 to -115, -90 to -70 in the P arm and from approximately +72 to +100 in the P' arm (Fig. 4A, lanes 1–3 and 7–9). The NCS footprint appears to affect specific residues in the P arm within the region of -145 to -120, -80 to -65, and the region +74 to +96 in the P' arm (Fig. 4B, lanes 1–3 and 7–9). When these results were mapped onto the P22 attP sequence the highly conserved 10-bp sequence TGGGACPuNPuN was noted at regions distal to the proposed core. In four out of five instances the residues indicated as Pu are adenines. We have designated these sites P1, P2, P3 and P'1 P'2 by analogy to the attB truncation experiments is the consensus sequence element (for summary see Fig. 6B). Further evidence that these conserved sequences represent the P22 Int arm type binding sites was obtained by methylation footprints. In five separate experiments methylation footprints were performed on top and bottom strand-labeled P22 attP, the results of which...
**Structure of the P22 att Site**

**A**

P22 ATTB

\[ \text{agctcagttctga} \]
\[ \text{CGGCGAT} \]
\[ \text{cctctatggct} \]
\[ \text{gcctctatttttg} \]
\[ \text{agctcagttctga} \]
\[ \text{CGGCGAT} \]
\[ \text{cctctatggct} \]
\[ \text{gcctctatttttg} \]

**B**

P22 ATT P

\[ \text{ctgattgctagtactttgcatcggtttgcaaggctttgcatgtctttcaa} \]
\[ \text{P1} \]
\[ \text{P2} \]

\[ \text{P3} \]
\[ \text{IHF} \]
\[ \text{C} \]
\[ \text{C'} \]

\[ \text{gaggtaaggaggttaaattCTTATM} \]
\[ \text{GACATAATTCG} \]
\[ \text{ctcttcttttgctctctttttga} \]

\[ \text{IHF} \]
\[ \text{P'1} \]
\[ \text{P'2} \]

\[ \text{ctccctttcagctcgaccgatttaaGACAATTCG} \]
\[ \text{caggtcatcttctttttgctctctttttga} \]

**FIG. 5. Summary of the nuclease protection of attP and attB by DNase I and NCS.** The relevant sequences of the attB (panel A) and the attP (panel B) and their corresponding coordinates are indicated. Regions protected against DNase I cleavage by Int and IHF are underlined. Bases protected against NCS are indicated by x. Specific enhancements (*) and protection (o) in methylation experiments are also shown. The consensus sequences for the Int core type and arm type binding and IHF binding are indicated by upper case letters. Sites of Int cleavage are indicated by open triangles. The boundaries of the minimal att sites are indicated by the dark vertical arrows.

---

**DISCUSSION**

Before discussing the rather striking differences between the P22 and \( \lambda \) att sites we shall consider the areas of commonality. The P22 att site, like that of \( \lambda \) and P2, has two classes of Int binding sites (core and arm type) separated by binding sites for the DNA-bending protein IHF (see Fig. 7). Efficient P22 recom-
bination occurs in the presence of E. coli IHF, whereas in its absence recombination is detectable but greatly depressed (data not shown). The P22 core type Int binding sites, which coincide with the core sequences predicted by Campbell (24), are identical at C' and B', whereas C and B, which straddle the boundary of partner homology, differ from each other and from C' and B' (Fig. 6A). At any given position within the core type binding sequence, at least three out of the four sites are identical, and the last three nucleotides are invariant (Fig. 6A). When the P22 core type consensus recognition sequence is compared with that of λ it is found that the sequences at positions 1, 5, and 9 are identical. The possible importance of these positions may be related to the observation that λ Int can cleave a λ attL which has a P22 consensus sequence inserted at the C' position, whereas λ Int cannot cleave an attL with a random sequence inserted at C' (37). When the P22 core type Int binding sequences are compared with those of P2, there is no significant identity; however, there is a shifted alignment in which five consecutive positions of the consensus sequences have the same bp, and four of these have the same alternative bp.

Comparison of the P22 Int core type and arm type consensus recognition sequences leaves little doubt that they are recognized by different protein domains as has been shown for A Int. Based on a comparison of the arm type Int binding sequences it appears that P22 Int may be more closely related to P2 than to λ Int. Seven out of the 8 bp in the P2 arm consensus match with the H' IHF site, are on the opposite face of the helix from the C' core type site. In contrast, the P22 P' arm has only two adjacent arm sites both with the same orientation, and both of which, along with the H' IHF site, are on the opposite face of the helix from the C' core type site. In P22, deletion of both of these arm type sites produces an attP incapable of recombination. On the other side of the core region, in the P' arm, λ contains two nonadjacent Int binding sites facing in opposite directions, whereas P22 has three Int binding sites, all facing in the same direction, two of which are adjacent. The adjacent P22 P arm sites are, like those in the P' arm, on the same face of the helix as each other but on the opposite face from the C core type site. Our data indicate that the attP must extend through the second P' site for recombination to occur (for summary see Fig. 5).

Based on a comparison of the arm type Int binding sequences it appears that P22 Int may be more closely related to P2 than to λ Int. Seven out of the 8 bp in the P2 arm consensus match P22, the one not matching being a single bp insertion at position 3 of the P2 sequence (15). Despite this similarity in consensus recognition sequence, the disposition of the four P2 arm type sites is still quite different from that of the five P22 arm type sites (Fig. 7).

In addition to these differences between the P22 and λ arm type Int binding sites, the number and relative positions of the IHF binding sites are also different (22). Although the basic theme of an IHF site between the core type and arm type Int sites is clearly the same in both systems, P22 has only two IHF sites, which asymmetrically flank the core, whereas λ has three, two of which are nearly equidistant from the center of the core region (43). The P2 attP, with only a single IHF binding site, represents an additional step in this continuum in the number of IHF sites (15). In P22, the helical face of the two IHF sites differs from each other, from that of the core type Int sites, and from the corresponding IHF sites in λ. In λ, all three IHF sites are required for integrative recombination and presumably shape the reactive higher order complex (44, 45).

The other DNA-bending protein that is required in the complexes for excisive recombination is the phage-encoded Xis protein. The P22 Xis protein has not been purified or studied, but we tentatively suggest that it might bind to two direct repeats

---

2. L. Smith-Mungo and A. Landy, unpublished results.
in the P arm (–118 to –81) in a manner similar to λ. These are the longest repeated sequences in the viral arms not accounted for by Int or IHF binding, and their position relative to core is roughly similar to that seen in λ. However, if these are the Xis type sites they too are not disposed relative to the Int arm type sites as they are in λ. And yet a third disposition is seen for the binding site of the Cox protein, which in P2 recombination is the functional analog of Xis (16). It is interesting to note that in all three phage, one arm of attP is longer than the other: a difference of approximately 60 bp in λ, 50 bp in P2, and 20 bp in P22. In λ and P2 the excision accessory proteins (Xis and Cox) bind in the longer attP arm, which is also the more likely arm for Xis in P22.

Given the complexity of these recombination pathways and the similarity in their multiprotein interactions, the above differences in the structure of the P22 and λ att sites are quite unexpected. We consider three possible interpretations of the differences. (a) There are significant differences in the higher order structure of each complex. Using the basic motif of binding proteins bringing together two kinds of sites that are transiently tethered by a bivalent protein, it is possible to draw out a large number of different structures. These structures could differ from each other in the path of the DNA within the complexes, the location and sign of DNA nodes, and/or the patterns of connectivity between the various arm and core type Int sites. (b) The complexes are well defined, but there is a considerable amount of flexibility in one or more of the proteins such that the same (or very similar) structures can be built from these different arrangements of binding sites. For example, if the hinge region joining the amino- and carboxyl-terminal domains of Int were very flexible (both rotationally and longitudinally), the observed differences in spacing, orientation, and even helical phasing might be readily accommodated within complexes whose overall structures are very similar. (c) There is no single specific structure that is competent for recombination pathways. Recombination in each system would be catalyzed by any one of several different structures that are quasi-equivalent in their capacity for recombination. According to this view, it is not at all surprising that the arrangement of protein binding sites in each system is different. This view further predicts that there will be many different arrangements seen within the complex Int family and even among the closely related lambdaid systems. The views described in b and c make the specific prediction that the arrangement of sites in one system would work in another and vice versa.

Insights that will allow us to distinguish among these possibilities are likely to come from a better structural understanding of individual complexes and also from comparisons among the different type II Int family members. Although these questions are obviously important to a satisfactory understanding of site-specific recombination, they also have interesting implications for more general questions about the structure of higher order protein-nucleic acid complexes that function in a wide variety of pathways.

Acknowledgments—We are greatly indebted to Martin Horvath, Jessica Rich, and Charles Leng for purified P22 Int protein; John Rossi for plasmid pJ1R1; Bob Weisberg for helpful comments on the manuscript; Tina Oliveira for assistance; Joan Boyes for preparation of the manuscript; and members of the Landy laboratory, especially Lina Moitoso de Vargas, for advice and helpful comments throughout this project.

REFERENCES